DESIGN AND CONTROL OF ACTIVE VISION BASED MECHANISMS FOR INTELLIGENT ROBOTS

Liwei Wu
email: liweiwu@ece.arizona.edu

Michael M. Marefat
email: marefat@ece.arizona.edu

Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721

ABSTRACT

In this paper, we propose a design of an active vision system for intelligent robot application purposes. The system has the degrees of freedom of pan, tilt, vergence, camera height adjustment and baseline adjustment with a hierarchical control system structure. Based on this vision system, we discuss two problems involved in the binocular gaze stabilization process. They are fixation point selection, vergence disparity extraction A hierarchical approach to determining point of fixation from potential gaze targets using evaluation function representing human visual behavior to outside stimuli is suggested. We also characterize different visual tasks in two cameras for vergence control purposes and phase-based method based on binarized images to extract vergence disparity for vergence control is presented. Control algorithm for vergence control is discussed.

I. Introduction

The advantages of active vision over passive vision in enabling the robot to explore its environment and then to adapt to the environment have been recognized by many researchers in active vision paradigm. As defined by Ruzena Bajcsy [1], active vision is a problem of intelligent control applied to data acquisition process depending on the goal or task of the process. It is able for the active vision system to improve its view point to overcome the inherent problem involved in passive vision that the sensor only takes in those percepts that randomly fall onto the sensors and thus, enlarges active vision based robot's adaptability to its environment.

From this definition we can elicit two points. The first is *what we want to see* (data acquisition depending on the goal or task of the process.). This is the problem of visual target selection. The second idea is *how to see the selected target* (intelligent control applied to data acquisition.). This involves determination of the position of the target and control of the vision system such that the target can be percepted. See Fig 1.1.
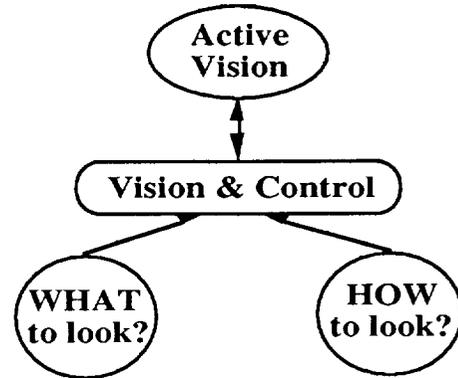

Fig 1.1   Concepts of an active vision system

Of importance to active vision is the gaze control strategy. Gaze control can be roughly partitioned into two categories [2]: Gaze Stabilization, which consists of controlling the available degrees of freedom for the active vision system such that clear images of interesting world point is maintained, and Gaze Change, which is motivated by the need to reduce computational complexity of visual tasks or to gaze at a new point that is taken into account for the visual tasks. This paper is concerned with problems in gaze stabilization.

From the point of view of binocular visual system, gaze stabilization means the visual axis of the two cameras point at the point of interest. The process of gazing at such a point is referred to as *fixating* and the point to be fixated at is known as *point of fixation*. Holding gaze at a selected target has several advantages in image processing. Gazing at the selected target means to capture the target in the part of the lens with highest resolution. This helps quantitative or qualitative visual performance. When the target is near the origin of an image, perspective projection model, which involves non-linearity, can be replaced by orthographic projection model that simplifies many computations. Since the fixation point has a stereoscopic disparity of zero, it is possible to use stereo algorithm that accepts limited range of disparity. This undoubtedly accelerates image processing. While the target is moving, fixating at it induces target "pop-out" [5] due to motion blur so that segmentation is much easier.

Basicly there are three problems involved in gaze stabilization, see Fig 1.2.
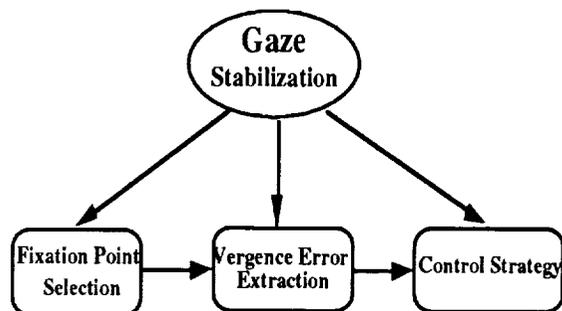
Fig 1.2   Three problems involved in gaze stabilization

The first problem in gaze stabilization is the determination of point of fixation FP. It is the first step in gaze stabilization. Gazing without a fixation point is ridiculous. The determination or selection of a point of fixation is to find the image coordinates of the fixation point's projection in the image plane in the presence of many alternatives based on some criteria. As active vision is a purposeful perception of visual targets, the selection of fixation point will depend on the goal of visual tasks.

The second problem is vergence disparity measurement. The process of two visual sensors' pan motion about their vertical axes in opposite direction to fixate at the selected point of fixation is called *vergence*. Since the optical axes are initially not pointing at a selected point of fixation, the vergence error must be derived so that they can be compensated for to ensure that both optical axis are keeping directed at the target.

The third problem is also the key point of general active vision research. An active vision system has mechanisms that can actively control camera parameters such as position, orientation, vergence, focus, aperture, etc. in response to the requirements of the task. Active vision system is, thus, not only a visual system but also a control system. The tasks of an active vision system are not only visual tasks but also control tasks. Therefore the third problem is the control strategy by which gaze stabilization can be fulfilled.

In this paper we are going to present the design of an active vision system and deal with these problems in binocular system's gaze stabilization with emphasis on fixation point selection and vergence disparity extraction. We introduce the concept of fixation point candidates (FPC's) in the image the cameras take and use evaluation functions to hierarchically determine the point of fixation among all the candidates. This approach is a mathematical representation of psychological results of human visual behavior so that our approach has a solid

theoretical foundation. Based on binarized images, we propose a method that robustly and efficiently extract vergence disparity signal, i.e., the vergence error. This error is the motivation of corresponding vergence control action of binocular system to ensure gaze stabilization. The method has certain advantages over existing approaches discussed in [3] and [5].

The paper is organized as follows. In the coming section, the design of our robot "head", i.e., the binocular active vision system will be presented followed in section III by the discussion of the approach to determining point of fixation, Then in section IV, vergence disparity extraction is discussed. The paper ends with conclusion in section VI.

II. A Binocular Active Vision System

1. Robot "Head"

To implement binocular active gaze stabilization, a particular apparatus is required to provide control over the acquisition of image data. From a mechanical perspective, a binocular active system has a mechanical structure which provides mechanisms for modifying the geometric or optical properties of two cameras mounted on it under computer control. One approach is the construction of a robot "head". The design of such a robot "head" includes the design of a mechanical structure on which the cameras are mounted, by which cameras positioning can be completed as well as the design of a control system that controls the cameras' movement and also camera's optical parameters (which is not going to be discussed in this paper.).

A robot "head" has at least the following degrees of freedom:
1) Pan, which is a rotation of the two cameras about a vertical axis passing the midpoint of the baseline;
2) Tilt, which is a rotation of the two cameras about a horizontal axis, e.g., the baseline;
3) Vergence, which is an antisymmetric rotation of each camera about a vertical axes passing through each camera.. See Fig 2.1 and Fig 2.2.

Several research groups have built some robotic heads subject to different design criteria and applications. As a matter of fact, different realization has its own advantages and disadvantages. As to active vision sensors, what is more important, it seems to us, is the ability to obtain accurate 3-D information and convenience implementation of gaze control. Baseline adjustment ability is added to the system in our "head" design apart from other degrees of freedom. Baseline adjustment is the change of distance between two vertical axes of the two cameras, assuming the vertical axis pass the focal point. It is considered to enhance the ability for accurate depth perception when the vision system is close to the
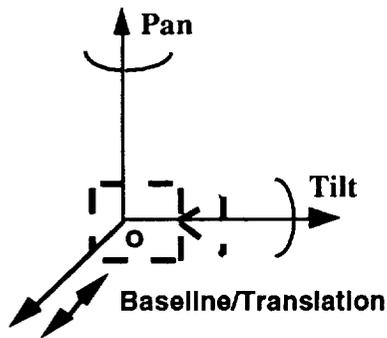
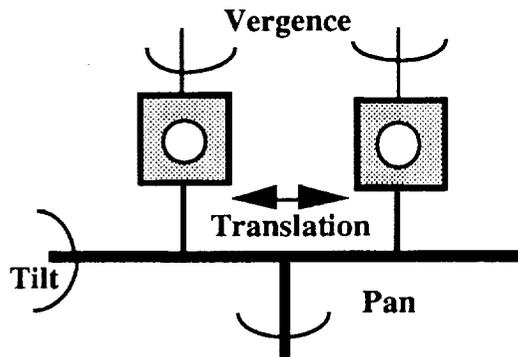**Fig 2.1 Pan, tilt motion of the robot head**

Fig 2.2 Degrees of freedom of the robot "head"

object, although the "baseline" of human visual system is fixed. Thus the cameras can translate along tilt axis. Note, this translation movement is antisymmetric. Secondly, the gaze ability of a binocular active vision system is the most significant advantage over any other types of vision system. We choose the structure as shown above in Fig 2.2 because this structure has several advantages over other possible designs in gaze control. In this design, the vergence angle and pan angle are controlled by separate motors (Pan angle is controlled by pan motor and vergence angle by vergence motors.) and are orthogonal -- either parameter can be altered without disturbing the other [3]. A mechanical advantage of this design is its simplicity: the compact mechanisms and fairly direct linkages facilitate rapid saccades change[3]. The structure of our robot "head" is depicted in Fig 2.3, where head's height adjustment ability is added in case of necessity.

2. "Head" on a Robot Arm

Although the "head" is provided with pan, tilt, vergence, and baseline adjustment motion abilities to change the cameras positioning and orientation to obtain various viewpoint for different tasks, there are still some vision problems in application that such a "head" cannot solve. Active vision system is not merely a vision system, it serves for action. It will cooperate with a robot arm to accomplish a specific task. In real application, the view could be obstructed when the robot arm is in close proximity to the object. Also, in CIM applications, the "head" may need to see the opposite face or a side face of a part. In such cases, we can clearly feel that more "degree of freedom" should be provided to the visual system, the head. This means that it is better to mount the vision head on the end-effector of a robot arm (See Fig 2.3). This configuration will offer maximum field of view for the cameras.
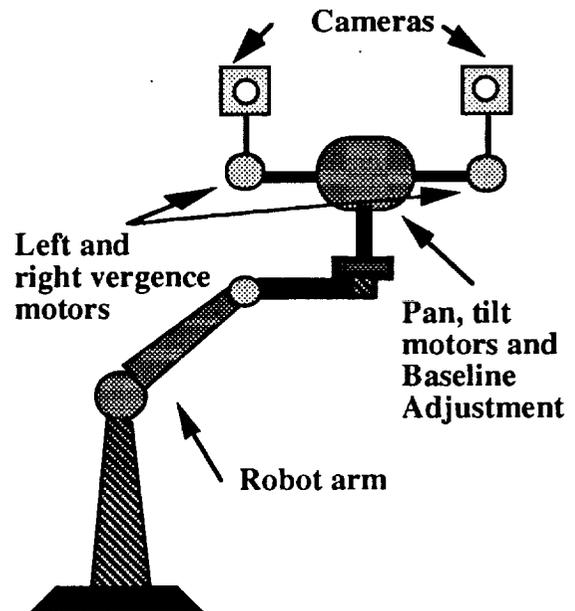

Fig 2.3 A "head" mounted on the end-effector of a robot arm

3. Robot Head's Control System Blocks

Each degree of freedom is actuated by a DC servo motor because of its easy controllability nature. The basic block diagram of the robot head's control system is shown in Fig 2.4. Each degree of freedom has its own local controller, which are coordinated by the robot head platform control block. The control block is interfaced to a host computer which is also the host computer of the whole active vision system. Control signals are synthesized in the host computer and sent to platform control block. The control block receives the command from the host, does kinematic calculation to get control signal for pan, tilt, vergence, or other motion control purposes, and then sends them to different local controllers to implement the control command from the host computer. The system forms a hierarchical control structure with three levels. The top level is the host. In the middle, platform sub-controller communicates with host and the bottom level local controllers as a coordinator. The bottom level local controllers are actual controllers for specific control task, such as pan, tilt, or vergence,etc.
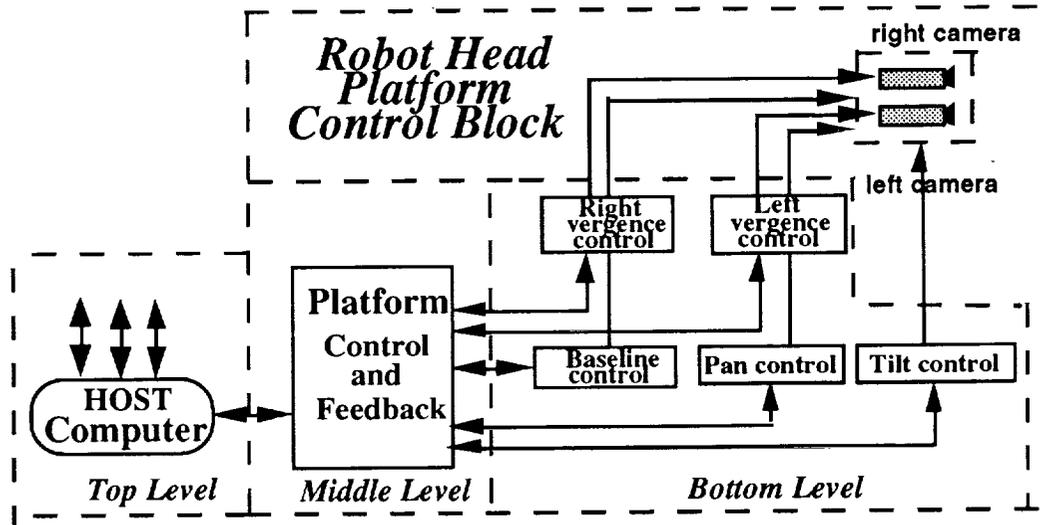
160

Fig 2.4 Robot head's control system block diagram

III. Determination of Point of Fixation

The general gaze stabilization problem is to maintain fixation on a (moving) visual target from a moving observer. In our case of binocular system, this means the axis of the two cameras point at the target. Thus, the positions of the projections of the target are at the origins of both image plane coordinate frames. Since the object the vision system "looks" is usually not a geometric point that has no volume the projection of the object in the image plane will not be a point but an area. Then the first question we encounter is "what part of the object should the cameras fixate at"?

1. Gaze Target and Its Selection

Gaze stabilization is closely related to visual tasks the system performs. The goal of present visual task determines what the system should gaze. This is true because focusing limited system resources on restricted region of the scene, or the most important region of a scene related to current visual task, is necessary from the point of view of cost and complexity considerations [2]. In this paper, we are not going to discuss the problem of "What I am going to look". This is related to "next look" problem and is beyond the scope of our discussion in this paper. What we discuss is the mechanism of gaze stabilization. The problem is "How I am going to look". This means we will tell the system what it should look. Once it is told what to look, it is system's responsibility to find the target and hold gaze at it.

Some human visual behaviors form our theoretical foundation of selection of gaze target. Human visual shifts when the visual systems are confront with a new stimulus. This stimulus will then become the new target the eyes are to fixate at. The shift is wholly dependent on the visual information and the result of the shift is to

bring the target onto the fovea, where resolution is highest. Psychological studies of human visual behavior to outside stimuli reveal that any detectable feature can be used to guide attentional shift, but color, high-contrast region and image area with high spatial frequency being important factors in visual search and that attention often shifts to areas of "information detail". In a simple case, when searching random 2-D polygonal form, eye fixation tends to concentrate on vertices. These two criteria are called Low-level visual stimuli criterion and High-level visual stimuli criterion, respectively [4].

Hence, the targets that the system may hold gaze at are corners/vertices or edge points in an image. We choose them as potential targets not only because of the fact that human visual attention often shifts to areas of "information detail [4] such as vertices, edges, and axis of symmetry, etc. but also, on the other hand, corners/vertices and edge points are the most "salient" features in a picture and are of extremely usefulness in vision research. Finally, corners/vertices and edge points are more "explicit" features than others that can be used for study of gaze stabilization. Generally speaking, we choose the most "salient" and "explicitly represented" feature in an object as our promising fixation target. Our fixation point selection is feature-based.

To select the point of fixation from among all the corners/vertices and edge points in a picture, we need a couple of tools. One is the approach to selecting it from all the regular corners/vertices and edge points. We use a hierarchical approach to find the gaze target, the fixation point. The other is the criterion used to help in the selection of point of fixation from potential candidates. The criterion will be represented in the form of evaluation function. Practically, when we are selecting our gaze target, these two tools are used combinedly. The process of gaze target selection is described in Fig 3.1.
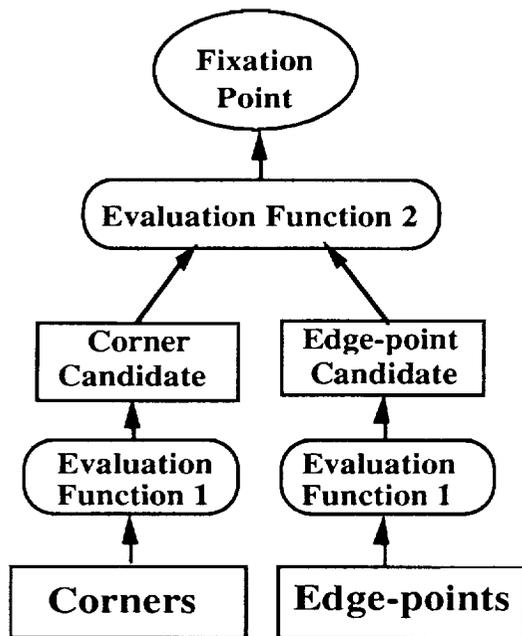
161

```
      ┌─────────────┐
      │  Fixation   │
      │   Point     │
      └─────────────┘
            ▲
      ┌─────────────────┐
      │ Evaluation Function 2 │
      └─────────────────┘
         ▲         ▲
   ┌──────────┐  ┌──────────┐
   │ Corner   │  │ Edge-point │
   │ Candidate│  │ Candidate  │
   └──────────┘  └──────────┘
       ▲             ▲
   ┌──────────┐  ┌──────────┐
   │ Evaluation│ │ Evaluation│
   │ Function 1│ │ Function 1│
   └──────────┘  └──────────┘
       ▲             ▲
   ┌──────────┐  ┌──────────┐
   │ Corners  │  │ Edge-points│
   └──────────┘  └──────────┘
```

Fig 3.1   A Hierarchical approach to the determination of fixation point

We first find all the corners/vertices and edge points in a picture. They form two separate groups. In each group, we use evaluation function to determine each group's possible gaze target (fixation point), which is called *fixation point candidate*. Between the two candidates, we again apply evaluation function (different from the former evaluation function in parameters, structure, and etc.) to find the gaze target, the fixation point. The detailed algorithm will be given in the later sections. In the following two sub-sections, we will first discuss detection of corners and special edge points in an image which form the mentioned candidate groups.

2. Corners and Special Edge Points

A. Related Work to Corner Detection

Corner detector as an image feature extractor has been discussed in many literature. Corners/vertices are important features of an object. They can be used for identification of an object in the scene, for stereoscopic matching, and displacement vector measuring [6]. In binocular system's gaze stabilization they are considered to be the most important fixation point candidates.

Since corner is also an edge point where curvature changes drastically, in the earlier approaches to detect a corner/vertex, image is first segmented and then the curvature of edges is computed. A corner/vertex is declared if the curvature at the point is greater than a pre-defined threshold and the point is also an edge point [8]. The other group of approaches of corner/vertex detection i.e., more recent approaches, is based directly on gray-level image. The effort was first made by Beaudet [7].

These methods measure the gradients of the image and use an operator to measure the "cornerness". These methods can be referred to [8][9][10][11], which are considered to be equivalent in nature [11].

An appropriate approach to corner detection for gaze stabilization application can be found in [18]. The approach searches for edges according to the gradient magnitude and direction to find a micro-intersection points, calculation of the distance from the intersection to the current point and keep of the minimum distance. After non-minimum suppression in the distance distribution map, all corners can be found. The algorithm is simple, reliable and noise insensitive and has good localization [18]. These are important reasons that this approach is chosen for our real-time corner-detection application.

B. Special Edge Points

Edge points are another class of "salient" features that can be considered as gaze target in gaze stabilization. Clearly, we are unable to search for edge candidate from among all the edge points since it is computationally much too expensive to do that. And in fact, it is not necessary to consider all the edge points. Physiological research tells us some other interesting properties of human visual behavior to outside stimuli. Proximity of Stimuli [4] states that for several potential targets in the visual field, the one which is closest to the fovea is more likely to be selected as a fixation target and Direction of Stimulus states that upward eye movement is preferred to downward movement. We may conclude that, for two potential new targets, the one that lies above and close to current origin of image frame is more likely to be selected as the next fixation target than the positionally lower and far target.

According to proximity stimuli criterion, we say only one specific edge point on an edge line segment that is closest to current origin of the image plane coordinate needs taking into account. An edge point which is closest to another point p_x (here it should be the origin) that does not lie on that edge line segment is the intersection point (p_e) of this edge line segment and the line which passes p_x and is perpendicular to that edge line segment, i.e., the foot of perpendicular. See Fig 3.2 (a).

In order to determine the edge point candidate, we draw vertical lines to each detected edge line segments from the origin of the image plane coordinate. The intersection points thus determined are of interest and from all these special edge points the edge point candidate will be selected.

But note, there are two cases in which the resulting intersection points will not be taken into account. The first case is that the intersection point is one of the
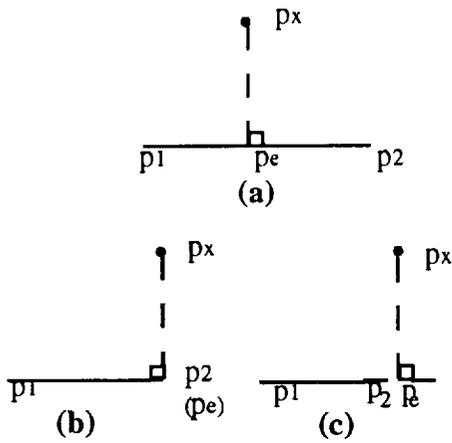
Fig 3.2 (a) Foot of perpendicular. (b) Intersection point is one of the end points. (c) Intersection point lies on the extended line of the edge line segment.

end points of the edge line segment, see Fig 3.2 (b). Since end points are also corners/vertices that have been considered, these intersection points are discarded. The second case is that the intersection point lies on the extended line of the edge line segment , see Fig 3.2 (c). Thus, the computed intersection point actually does not exist. These points also can not be considered. We propose a simple method to detect if a computed intersection point is on the extended line.

In the case of Fig 3.2 (a), point p_e lies on the line segment. we have:

$$\overline{p_1p_e} + \overline{p_ep_2} = \overline{p_1p_2} \qquad (3.1)$$

In Fig 3.2 (c) where intersection point lies on the extended line, we have:

$$\overline{p_1p_e} + \overline{p_ep_2} > \overline{p_1p_2} \qquad (3.2)$$

When (3.2) holds, we should discard the computed intersection point p_e

C. Fixation Point Candidates Determination

Now, all the corners/vertices detected and edge points that are computed form two groups. We are going to determine the fixation point candidate (FPC's) in each group. The approach to determine the FPC's is based on the psychological studies conclusions on human visual behavior. An evaluation function which represents both proximity of stimulus and direction of stimulus criteria is formulated to aid in the decision making of fixation point candidate selection. This first evaluation function takes the form of:

$$FPC_i = min \; \{\alpha X_i^a, X_i^b\} \qquad (3.3)$$

where X denotes either a corner (then $X \stackrel{\Delta}{=} C$) or an edge point (then $X \stackrel{\Delta}{=} E$), a and b represent those points that are positionally above or below the current origin of the image plane coordinate frame. X_i (i = 1, 2, ..., j, the number of corners detected or special edge points that are

computed.) is computed as Cartesian distance between the point and the origin and thus is:

$$X_i = \sqrt{p_x^2 + p_y^2} \qquad (3.4)$$

where p_x and p_y are the coordinate values of the point being considered.

α is a constant between 0 and 1, i.e., $0 < \alpha \leq 1$. This weight represents the criterion of direction of stimulus.

Then the points, a corner and an edge point, will be selected as corner fixation point candidate and edge point fixation point candidate in each group if they have the minimal values of FPC_i in each group. The two selected candidates have the distances C_{FPC} and E_{FPC} from the origin, respectively.

D. Fixation Point Determination

Fixation point will now be determined between the two candidates. The criteria for the selection is also to apply mathematical representation of psychological results in the form of evaluation function. The second evaluation function for the final fixation point selection is:

$$FP = sgn \; \{[b^*C_{FPC} - E_{FPC}] + [D(C_{FPC}) - D(E_{FPC})]\} \qquad (3.5)$$

where sgn(.) is a sign function and D(·) is the measure of the dimension of the point being considered. If the point lies on one of the coordinate axes, its dimension is 1, otherwise the dimension is 2. This is a measure for control implementation. Larger dimension means more control actions will be concerned.

β is a constant and $0 < \beta \leq 1$. This weight used here represents the intention that corner is more preferred to be selected than edge point candidates due to High-level visual stimuli criterion.

Thus, if FP > 0, which means either the distance and dimension of the corner candidate are greater than those of the edge candidate or much control will be concerned though the distance of the corner candidate is slightly shorter than that of the edge candidate, then the edge point candidate will finally be selected as point of fixation.

If FP < 0, which means the opposite situation to the above discussion, then the corner candidate will finally be selected as point of fixation.

We may derive from the above discussion that the determination of fixation point not only depends on the features themselves but also the weights we select, i.e., α and β. In some sense, the selection of α and β has important influence on decision making on fixation point selection. We propose that $\alpha = 0.9 \sim 0.95$ and $\beta = 0.95 \sim 0.99$.

The algorithm for determination of the point of fixation is given below:

1) For each corner or special edge point in each group, calculate its distance X_i from the local origin using (3.4),
2) Determine the candidate for point of fixation in each group using evaluation function 1 represented by (3.3),
3) Determine the point of fixation using evaluation function 2 represented by (3.5),
4) Get the coordinates of the selected point of fixation: (x_{FPL}, y_{FPL}).

IV. Vergence Disparity Measurement

1. Problem Description

As mentioned before, gaze stabilization in binocular system means pointing the two optical axes of two cameras to the selected fixation point. Thus, the positions of the projection of the fixation point are at the origins of the two image planes. The process of realizing fixation is called *vergence*. A straightforward and easy way to do this is to select the fixation point in different cameras separately and control the parameters of the degrees of freedom available to each camera such that the fixation point projects onto each origin of the image planes coordinate frame. However, this method is not reliable. The reason is that if fixation point is selected separately in two cameras, we are unable to say that the two cameras will select the same point because geometrically the initial positions of projection of the object in two images are quite different. The approach proposed does not guarantee global determination (which means determination of position of a visual target in two images.) of the position of fixation point. This results in non-fixation in real application.

Then , what is a reliable method? Remember the vergence system is also a control system. From the view point of a closed-loop control system, the measure of the difference, or error, between the desired input and the actual output is important since control signal is synthesized based on this error signal [22]. Back to our vergence control, let's ask: "What is the error signal involved in vergence control"? We know that fixation point has a stereoscopic disparity of zero. This is a "salient" feature of fixation. To achieve fixation means to obtain zero disparity between two images. If the disparities between the two cameras are zero, we are sure that the two cameras are fixating at the same point. So to compensate the disparity between two images is a direct and reliable approach to realizing fixation.

If we accept this conclusion and try to find the disparities, one of the images in the two cameras should be considered as the reference image. If the image of the left camera is chosen as reference image, we say the left camera is the dominant camera [4]. Th a, the task of fixation point selection only affects the dominant camera. The tasks involved in the dominant camera and its sub-control system are:

1.(optional) Tracking if the target is in motion with respect to the dominant camera,
2. Fixation point selection, and
3. Control of degrees of freedom to keep the optical axis directed to the fixation point.

Now we can consider the image in the other camera, the non-dominant camera, as the "output" of the vergence system. Then, the difference or the disparity between two images, are the *error* signal of a vergence system. So we need to control the parameters of the degrees of freedom available to the non-dominant camera such that the disparity is compensated. When vergence control results in zero-disparity, we believe that the two cameras fixate at the same target. Therefore, tasks involved in non-dominant camera and its sub-control system are:

1. Vergence disparities extraction, and
2. Disparity compensation (vergence control process). Refer to Fig 4.1

There are a lot of algorithms that deal with disparities [16][17][18]. They are usually used to obtain a depth map. In disparity estimation for vergence control, what we need is an "overall" disparity estimation --- the disparity between the images. The whole image could be regarded as a single "big point". Our approach is Fourier phase-based approach. It is motivated by the Fourier translation property that a translation in spatial domain will result a translation in frequency domain that is direct proportional to spatial translation. When disparity exists in two images that are taken at the same time but in
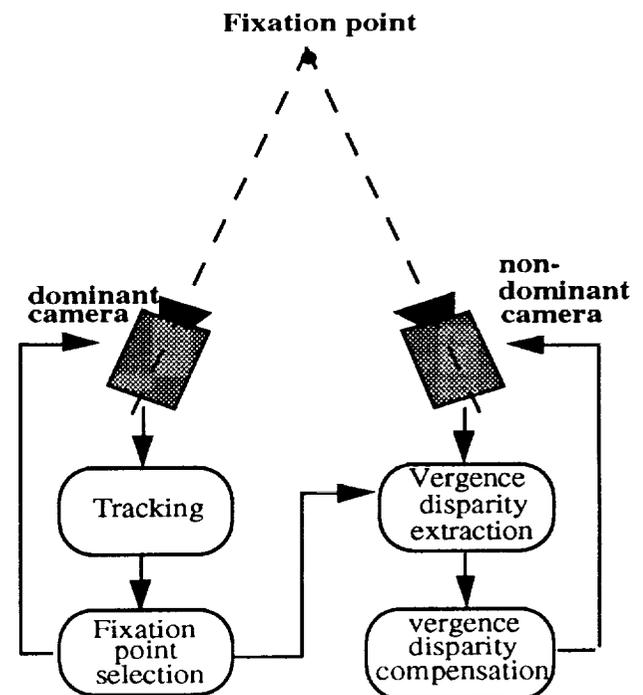

Fig 4.1 Different tasks in left and right camera for fixation

164

different cameras, we can regard the two images as taken consecutively in one camera and the disparity is due to the translation of the object. Thus, by calculating the phase difference of two "consecutive" image, we are able to determine the translation of the object in two consecutive images and then the actual disparities can be determined. Our approach is similar to [13] in that the two methods both use phase difference as a measure of disparity. But in [13], local disparities are important and this is why a local filter (Gabor filter) is involved since its goal is to obtain a depth map. In our approach, since we are only interested in "overall" disparity, the complicated gray-level images are used as binary images and treated as a single "large" point. Any local analysis is not necessary. Therefore, our approach is more suitable to vergence control.

The advantages of our approach over the existing approaches [3][5] for vergence control are:
1. We simplify the image processing --- gray-level images are used as binary images. The ideal and the seemingly unrealistic assumption (shifted version) becomes true in our approach.
2. The disparity is obtained directly as a function of the image property (Here only the contour is important.). It avoids the disadvantages contained in peak-finding method [12].
3. This approach is a robust estimation of disparity. Local occlusions and local intensity changes will not affect the "overall" disparity estimation.
4. It is simpler in that only phases are calculated. The computationally more expensive process of spectrum calculation is avoided while in [3][5] peaks are found in the spectrum analysis. Thus, presented approach is more suitable to real time application.

2. Vergence Disparity Measurement Based on Fourier Phase Difference

It is known that the Fourier phase difference between two consecutive images provides all the information required to obtain the relative displacement vector[15]. The most important advantage of using complex phase of Fourier transform in objection position detection is that a translation in the spatial domain directly corresponds to a phase shift in the spatial frequency domain. When an object is completely inside the image window, the relationship between position and fundamental frequency complex phase is linear [17][15]. More explicitly, the position and the fundamental frequency complex phase satisfy the following equation:

$$\Delta position = \frac{window\_size}{2\pi} * \Delta phase \qquad (4.1)$$

This equation can be directly obtained from the translation property of the Fourier transform represented by [24]:

$$f(x-x_0, y-y_0) \Leftrightarrow F(u, v)exp[-j2\pi(ux_0 + vy_0)/N] \qquad (4.2)$$

where we only consider fundamental frequency $(u = v = 1)$ and N is the window size.

If we regard the right image $R(x, y)$ as an image that is taken in the left camera right after the image $L(x, y)$ is taken and contribute the disparity to the shifts of the movement of the object with respect to the left camera, then, by calculating the fundamental frequency phase change in these two "consecutive" images, we are able to determine the disparity x_d and y_d. Once the disparities are determined, mapping them into vergence control system's reference input is not difficult.

It should be pointed out that the method introduced needs 2-D Fourier transform computation. One way to achieve faster processing is to use Fourier phase in conjunction with projection concept [15]. The use of projection is important because, in this way, it is possible to achieve 1-D processing and disparity x_d and y_d can be directly and separately obtained.

The projection of $F(x, y)$ along y-direction onto x-axis perpendicular to y-axis is defined by [15]

$$F_y(x) = \int F(x, y) \, dy \qquad (4.3)$$

Similarly, we have projection of $F(x, y)$ along x-direction onto y-axis:

$$F_x(y) = \int F(x, y) \, dx \qquad (4.4)$$

If we consider digital images, the integration should be represented as summation. Thus, equations (3.3) and (3.4) becomes:

$$F_j(i) = \sum_{j=0}^{h} F(i, j) \qquad (4.5)$$

$$F_i(j) = \sum_{i=0}^{w} F(i, j) \qquad (4.6)$$

where $h \times w$ is the window size and $F(i,j)$ is quantized from $F(x, y)$.

The algorithm below describes the procedure for vergence disparity extraction.

1. Determine an appropriate sized window such that the object is entirely within the window.
2. Get the projections of both images along x-direction and y-direction using:

$$L(i) = \sum_{j=0}^{h} L(i, j), \quad L(j) = \sum_{i=0}^{w} L(i, j) \qquad (4.7)$$

$$R(i) = \sum_{j=0}^{h} R(i, j), \quad R(j) = \sum_{i=0}^{w} R(i, j) \qquad (4.8)$$

3. Calculate their vertical and horizontal phases, which will be denoted by θ_L^i, θ_L^j, θ_R^i and θ_R^j, respectively.
4. The difference between the two pairs of phases will be

$$\Delta\theta^i = \theta_R^i - \theta_L^i \qquad (4.9)$$

$$\Delta\theta^j = \theta_R^j - \theta_L^j \qquad (4.10)$$

indicate the vertical and horizontal disparities according to (4.1).

$$x_d \int \frac{h}{2\pi} \int \Delta\theta^i \qquad (4.11)$$

165

$$y_d = \frac{w}{2\pi} * \Delta\theta^j \qquad (4.12)$$

As we have known the coordinates of the point of fixation in the left image are x_{FPL}, y_{FPL} and the disparity is (x_d, y_d), the coordinates of the point of fixation in the right camera will be (x_{FPR}, y_{FPR}), which satisfy $x_{FPR} = x_{FPL} + x_d$ and $y_{FPR} = y_{FPL} + y_d$ and which will be the reference input to vergence servo system after kinematic transform.

V. Control Issues

The x_{REF} and y_{REF} are in terms of pixels. They should be transformed to other two values in terms of pan degrees or vergence degrees or tilt degrees, etc., through kinematic calculation since this is the only form the local controller can accept. As mentioned before, each degree of freedom has its own local controller., which are coordinated by the robot head platform control block. The presently implemented control algorithm is PD algorithm, i.e., the output of the controller is proportional to the error between reference input and system real output and the derivative of the error. This is a typical implementation for DC motor drive system and can be mathematically represented as:

$$u(t) = k_p * e(t) + k_d * \dot{e}(t) \qquad (5.1)$$

where $e(t)$ is the error between reference input $r_i(t)$ and system's real output $y(t)$, i.e.,

$$e(t) = r_i(t) - y(t) \qquad (5.2)$$

Different choices of the two parameters of the PD controller, k_d and k_p, will result different output response. the larger the k_p, the smaller the steady error but the larger the overshoot. The larger the k_d, the more sensitive the system, either speeding the response or resulting oscillation. So the two parameters are empirically selected such that the step response of the system is slightly under-damped to achieve fast response with small overshoot. The simulation of one of the controller's output is depicted in Fig 5.1.

VI. Conclusions

The design of an active vision system is given with emphasis on the ability to obtain accurate 3-D information and on the convenience for gaze control. Based on this design we discussed three problems involved in binocular system's gaze stabilization process.

In fixation point selection, we argued what kind of features can be chosen as fixation point candidates. In this paper, we select corner/edge-point as salient feature for fixation purposes. Studies in human visual behavior provide us with theoretical foundation based on which evaluation functions are formed to determine fixation point hierarchically from between the candidates. We should point out that appropriate target for fixation are

chosen according to visual tasks the system is performing. Gaze control at the higher level can be
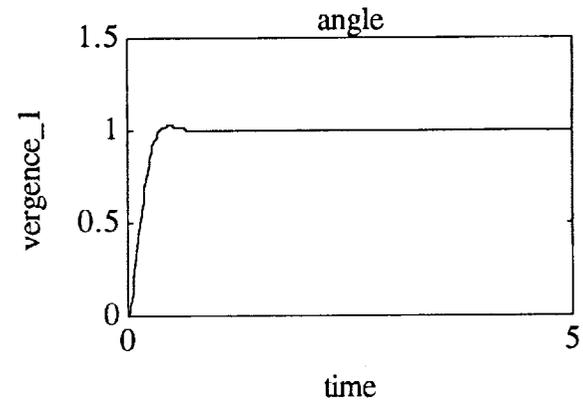
(a)
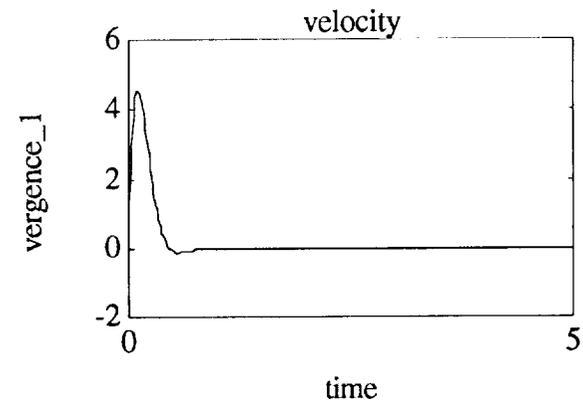
(b)

Fig 5.1 (a) Vergence servo output with small overshoot under step input. (b) The velocity of the output.

viewed as a resource management problem [3]. This is beyond the scope of this paper and is not taken into account. Here, we assume that corner/edge-point could be our appropriate target for fixation.

We characterized different tasks in left and right cameras for vergence control and used phase-based method to measure vergence error based on binarized images. This approach can robustly and efficiently extracts vergence disparities.

And in the last section we discussed some properties of the local controller based on PD algorithm.

VII. References

1. Ruzena Bajcsy, "Active Perception", Proceedings of the IEEE, Vol. 76, No. 8, August, 1988.1

2. University of Chicago, "Promising Directions in Active Vision", University of Chicago Techinical Report CS 91-27, November, 1991.

3. Thomas J. Olson and David J. Coombs, "Real-Time Vergence Control for Binocular Robots", Int. J. Computer Vision, 7:1, 1991.

4. A. Lynn Abbott, "A Survey of Selective Fixation Control for Machine Vision", IEEE Control Systems, August 1992.

5. David J. Coombs and Christopher M. Brown, "Cooperative Gaze Holding in Binocular Vision", IEEE Control Systems, June 1991 .

6. Rachid Deriche and Gerard Giraudon, "A Computational Approach for Corner And Vertex Detection", Int. J. Computer Vision, 10:2, 1993.

7. P.R. Beaudet, "Rotational Invariant Image Operator", *Int. J. Conf.* Pattern Recognition, pp. 579-583, 1978.

8. O.A. Zuniga and R. Haralick, "Corner Detection Using the Facet Model", IEEE CVPR Conference, pp. 30-37, 1983.

9. L. Kitchen and K. Rosenfeld, "Gray-level Corner Detection", Pattern Recognition Letter, Vol. 1, pp. 95-102, 1982.

10. H.H. Nagel, "Displacement vectors from Second-order Variations", Computer Vision, Graphics, and Image Processing, Vol. 21, pp. 85-117, 1983.

11. Mubarak A. Shah and Ramesh Jain, "Detecting Time-Varying Corners", Computer Vision, Graphics, and Image Processing Vol. 28, pp. 345-355, 1984.

12. Michael Jekin and John K. Tsotsos, "Techniques for Disparity Measurement", CVGIP: Image Understanding, Vol. 53, No. 1,, January 1991.

13. T.D. Sanger, "Stereo Disparity Conputation Using Gabor Filter", Biological Cybernetics, Vol 59, pp. 405 - 418, 1988.

14. D. De Vleeschauwer, "An Intensity-Based Coarse-to-Fine Approach to Reliably Measure Binocular Disparity", CVGIP: Image Understanding, Vol. 57, No. 2, March, 1993.

15. Mansur Kabuka, *et al,* "Robot Vision Tracking System", IEEE Transactions on Industrial Electronics, Vol. 35, No. 1, February, 1988.

16. Katsuhiko Ogata, Modern Control Engineering, Englewood Cliffs, 1970.

17. Gonzalez and Wintz, Digital Image Processing, Second Edition, Addison Wesley, 1987.

18 Huang Zhen Hua and Yu Qian, "A Direct Corner Detecting Algorithm", IEEE CVPR, 1986.